MAX-PLANCK-INSTITUT FÜR WISSENSCHAFTSGESCHICHTE
Max Planck Institute for the History of Science

2015

PREPRINT 474

N. Bulatovic, B. Saquet, M. Schlender,
D. Wintergrün, F. Sander

Digital Scrapbook – can we enable interlinked and recursive knowledge equilibrium?

Digital Scrapbook – can we enable interlinked and recursive knowledge equilibrium?

N. Bulatovic[1], B. Saquet[1], M. Schlender[1], D. Wintergrün[2], F. Sander[1]

1) Max Planck Digital Library, Munich, Germany
2) Max Planck Institute for the History of Science, Berlin, Germany

We investigate possible tools and approaches to develop a Digital Scrapbook, a virtual research environment inspired by the recursive nature of research for scholars where they can combine web and own resources into a new scholarly edition readily enabled for Open Access. Web resources are interlinked in the digital scrapbook by content capture and detail selection, rather than sole bookmark or link to resource URL, along with necessary accompanying metadata. We analyse several open source and commercial tools, with special focus on a Scrapbook-X Firefox Add-On, in order to match to desired Digital Scrapbook features. We further address the wider requirement context for development of such Digital Scrapbook environment, discussing both technical and user experience dimensions. We conclude with a recommendation on how to approach the development and operation of a Digital Scrapbook environment and provide rough estimates of the required resources.

Introduction

Traditional approaches to historical humanities research such as collection of sources, reproduction of necessary excerpts, analysis and then writing a paper have undergone significant change through innovations in areas of computer science, knowledge management and humanities scholarship itself. Technology development enabled rapid digitization of analogue sources and their reuse in a wide variety of research contexts. Moreover, new sources are digitally born, interconnected, represented and published on the Internet in novel and unprecedented forms, thus extending the research method and its potential, as well as the editorial aspects of representation, analysis and dissemination of research outputs (Renn, Beyond editions: historical sources in the digital age, 2014). These are no longer seen as simple research papers, rather as multimedia, interlinked, interactive content, combined and constituted from parts of different sources available on the Web, enriched by annotations and semantically related, thus contributing to the notion of the Epistemic Web (Malcolm D. & Renn, 2012) - “a sustainable ecology of knowledge, affording a place for established knowledge and creating space for new knowledge” (Malcolm D. & Renn, 2012, p. 835), realized through rich federated documents, openly accessible to everyone.

The Scrapbook approach

The “Digital scrapbook” is envisioned as a virtual research environment (Renn, Beyond editions: historical sources in the digital age, 2014, pp. 27-28) which enables for a humanities scholar to combine, annotate and manipulate content from various digital and digitally accessible sources on the Web into a single “document”. The idea of the “document” as a product of the digital scrapbook is not to be conceived in a traditional sense, such as a PDF, text or a rich text document – it is rather a

new combined content resource using parts of selected digital representations of original sources, maintaining at the same time the link to the source itself and the excerpt selected by scholar. Scrapbook sources are either content directly authored by the scholar, or a derivative (digital view, digital excerpt) of an existing foreign content.

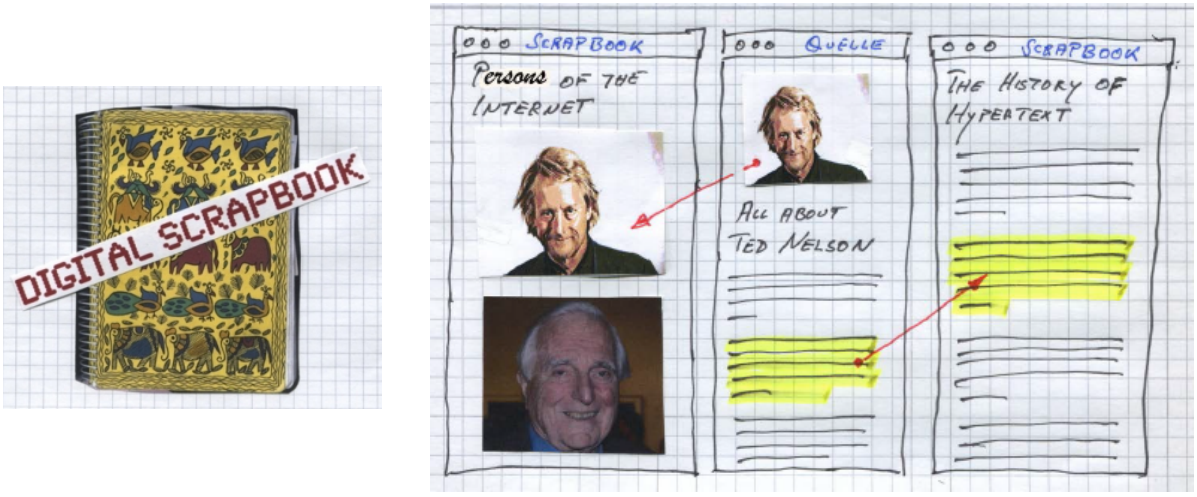


Figure 1: Digital scrapbook wireframes (Renn, Fiesole Collection Development Retreats, 2015, p. 26)

The scrapbook paradigm thus contributes to the idea of open, accessible and extendable knowledge and knowledge equilibration by implementing functionalities which support several dimensions:

- Permanent access – content referenced or reused into a scrapbook document is always available with the scrapbook document, thus avoiding a priori the link rot[1] problem
- Recursion and reusable content - all sources can be used and combined in several scrapbook documents, as well as a scrapbook document itself can be combined with other sources in a new scrapbook document, thus supporting the recursive nature of scholarship, reusing achieved results as means for further research
- Traceability, reproducibility and connectivity – content provided into a scrapbook document can be traced directly to the external or internal sources/knowledge it references, it is traceable and linked to the original external representation, assuming the latter is still available[2]
- Interactivity – a scrapbook document is openly accessible, can be searched and used in more powerful ways (e.g. fulltext search, linked with natural language processing i.e. NLP[3], data or computational services).
- Provenance – content combined in a scrapbook document should have complete provenance in order to preserve authorship, minimize legal and ethical issues, thus facilitating collaboration over content creation and sharing
- Discipline- and research question agnostic – despite it may not have a built-in basis to integrate methods and tools used in a particular discipline or necessary to help answering

[1] see https://en.wikipedia.org/wiki/Link_rot
[2] This is not a problem which Scrapbook can solve independently
[3] For recent review of NLP Technologies, see (Cambria & B.White, 2014)

specific research question, it can work well with numerous representations of sources from variety of domains (even though intended for humanities scholarship in the first place)

Is there a right tool for the job?

To define and understand better the functionalities of a potential Scrapbook environment, the MPDL Team[4] had made preliminary evaluation of several platforms, tools and javascript libraries, from open source to commercial tools. Preselection of tools was based on several criteria, in the first place, if the tool is free to use, if/how the tool can be extended, if it supports or is expandable for linking of resources, potentially as Linked Data (Berners-Lee, 2006)[5], whether the tool has already been used in similar scenario etc. Evaluation focus was reduced to the several tools[6], with a goal, to identify the tool which satisfies most of the Scrapbook requirements "out of the box" and can serve as a guideline for further implementation:

- Laverna (https://laverna.cc) is a personal open source note taking application, written in JavaScript, thus it requires no installation or registration. Notes can be made available via cloud storage, thus synchronized across several devices. The application stores all notes in local browser storage, such as indexedDB or localStorage. No user management, no multiple sharing, all notes are exclusively private. Notes' taking is simple and supports the Markdown[7] syntax. Laverna notes support links to external sources and uploading of images.
- ScrapBook-X (https://addons.mozilla.org/en-us/firefox/addon/scrapbook-x/) is a Firefox browser plugin for bookmarking/selecting/saving content from various web pages and creating a local collection of resources. Online resources can be either bookmarked, or captured locally. In addition it provides editing features for locally fetched resources, as well as the possibility to organize them into a collection or, to combine multiple resources into a single new document. The Scrapbook-X has been derived from the original development of similar Firefox plugins: ScrapBook by Gomita[8], and ScrapBook Plus by Haselnuss[9].
- TiddlyWiki (http://tiddlywiki.com/) is[10] a non-linear notebook to capture, organize and share complex information. There is a customized version of this application for scholars, to create and enter related persons, notes, ideas, concepts, theories in a more organized manner. It supports better organization of content in a personal Wiki, see the William Shakespeare[11] example.
- TagSpaces (http://www.tagspaces.org/) is a personal data manager, it helps organizing files by means of tags on every platform. It also provides rich notes taking features with a WYSIWIG editor. All data are stored locally, and can be shared via cloud storage such as DropBox[12],

[4] Max Planck Digital Library. https://mpdl.mpg.de
[5] See http://www.w3.org/DesignIssues/LinkedData.html
[6] The evaluation was not intended to find the best or the most appropriate tool
[7] See https://en.wikipedia.org/wiki/Markdown
[8] See https://addons.mozilla.org/en-US/firefox/addon/scrapbook/ and https://addons.mozilla.org/en-us/firefox/user/gomita/ for more information
[9] See https://addons.mozilla.org/en-US/firefox/addon/scrapbook-plus/ and https://addons.mozilla.org/en-US/firefox/user/haselnuss/
[10] As claimed from TiddlyWiki developers
[11] See http://tw5.scholars.tiddlyspot.com/#William%20Shakespeare
[12] See https://www.dropbox.com/

OwnCloud[13] etc. With its webclipper extension, it allows for clipping content from Web pages, tagging it and storing it locally.

- Pundit (http://thepund.it/) is a web annotation tool to annotate resources published on the Web. Besides Web page annotations it supports image annotations as well. In addition, it allows users to comment, bookmark or tag web pages, but also to create semantically structured data while annotating and browsing the Web. Pundit can be integrated with NLP[14] services and external knowledge bases such as DBPedia[15], thus pulling knowledge from existing Web of Data and Linked Open Data (LOD)[16], as well as enriching it through user-generated knowledge and annotations.
- Evernote (https://evernote.com) is a note taking and organizer cloud service, allowing its users to organize notes, ideas, tasks etc. and discuss and share with other users, as well as synchronize user content across different devices. In addition, it allows clipping content the Web resources. In basic mode it supports reduced set of features which can be used free of charge, with limited storage capacity (60MB). Features such as accessing notes in offline modus, generating presentations, identifying related content, integrate with email, PDF or Office documents are available against payment[17].
- OneNote (https://www.onenote.com/) is a note taking and organizer cloud service – functionally is similar to the Evernote service and is easily integrable with other Microsoft products. It allows for clipping and annotating content from anywhere on the Web, as well as linking notes to both web and local resources. Notes can be stored locally or synchronized across devices via various cloud storage services. In addition if offers many interesting features such as collaborative editing, translations, publishing to a Wordpress blog etc (Maurer, 2014).
- Annotator.js (http://annotatorjs.org/) is an open source Javascript library for adding annotation functionality to web pages, via plug-ins it can support image annotation, working with offline documents saved locally, sharing text and video annotations using social networks or email, and providing rich media and rich text annotations using the Tinymce[18] editor.
- The Note Taking Environment (NTE, http://www.cendari.eu/virtual-research-environment/) from CENDARI[19] Project is a virtual workbench for note taking and creation of Archival Research Guides (Pawliczek, 2014) that enables linking of data from various sources, integrates (Edmond, Bulatovic, & O'Connor, 2015) with Elastic search engine[20] for CENDARI sources, Medieval sources meta-search engine (TRAME[21]), an NLP Service for entity recognition, as well as resolution of entities against DBPedia and the internal CENDARI knowledge base. Notes are saved in an RDFa[22] format and stored locally, but as well sent to a common data repository based on the CKAN[23] software. The NTE can be extended to work together with the Pundit for building the CENDARI knowledge base.

[13] See https://owncloud.org/
[14] Natural Language Processing Services such as AlchemyAPI (http://www.alchemyapi.com/), AYLIEN (http://aylien.com/text-api) etc.
[15] See http://wiki.dbpedia.org/
[16] See (Berners-Lee, 2006) and http://www.w3.org/2013/data/
[17] See https://evernote.com/pricing/?offer=www_menu
[18] See http://www.tinymce.com/
[19] See http://www.cendari.eu/
[20] See https://www.elastic.co/
[21] See http://git-trame.fefonlus.it/
[22] See https://rdfa.info/
[23] See http://ckan.org/

Most important criteria that affected the choice for detailed analysis of the library, tool or service were as follows:

- It should provide open source codebase
- It should be developed with community of official standards in sight
- It should enable data to be stored locally
- It should enable sharing of data
- It should be simple and enable quick start for end users, no matter the underlying infrastructure
- It should be easily extensible
- It should be easily combined with other tools, primarily web browsers

While all closely observed components provide many features relevant for the Scrapbook approach, none provides all features by nature. For example, while OneNote and Evernote are pretty feature-reach, they do not fulfil the criteria of provisioned open source codebase, on the other hand side, they may not be quite easy to start with efficiently. TagSpaces is still a file manager and it cannot be used "as-is" for the Scrapbook purpose, but should be extended or integrated with additional components. Despite, it is great idea and an open source solution, whose developments are interesting to monitor: it works cross-platforms and its availability both as desktop and mobile app is contributing to this impression. Pundit is excellent at annotating images, texts and web pages, and is integrable with wide variety of knowledge bases, allows building own knowledge base, thus stands very close to the "prosumer" idea (Malcolm D. & Renn, 2012, p. 834) – however, it is still under development and is not easy[24] to use, there is some more work which needs to be done in this direction. Laverna on the other hand side is too simple, while TiddlyWiki is basically a Wiki solution, intended rather as a personal digital organizer, with options for sharing. The Annotator.js is just a javascript library, and can be integrated with other components to fulfil Scrapbook requirements. The CENDARI NTE is still under development and is not yet a finished product. Most of the tools, except Pundit, NTE, Evernote and Onenote do not provide user management and authorization features. The Scrapbook-X has been analysed in detail, with the aim to understand its potential, as it seemed to be the closest feature-related component with the Scrapbook paradigm, furthermore it is implemented as a Firefox browser plug-in i.e. at can be smoothly integrated in a web-based research and discovery of resources with a standard browser application.

Scrapbook-X

The Scapbook-X is a Mozilla Firefox browser plug-in, which helps saving web pages and organizing the web pages within a collection. The plug-in has been developed by Danny Lin[25] and has been very popular, along with related Scrapbook and Scrapbook Plus Firefox users.

Name	**Scrapbook-X**
Related developments	Firefox Plug-ins: Scrapbook, Scrapbook Plus

[24] Rather for expert and semantic data literates
[25] See https://addons.mozilla.org/en-us/firefox/user/danny0838/

	Differences: https://github.com/danny0838/firefox-scrapbook/wiki/Diffs
User documentation	https://github.com/danny0838/firefox-scrapbook/wiki/Intro
Source code	https://github.com/danny0838/firefox-scrapbook/
Developer	Danny Lin, https://addons.mozilla.org/en-us/firefox/user/danny0838/
Add-ons	https://github.com/danny0838/firefox-scrapbook/wiki/Addons MAF Reader: https://addons.mozilla.org/en-US/firefox/addon/mozilla-archive-format/ MAF Creator: https://github.com/danny0838/firefox-scrapbook-maf-creator or https://addons.mozilla.org/en-US/firefox/addon/scrapbook-maf-creator/

Table 1: Scrapbook-X information at a glance

Basic features which Scrapbook-X provides are:

- Web page capture – capturing the web page displayed on screen without losing any styles. Recording with it the capture time and the source URL for a later reference.
- Partial content capture – Users can select only part of currently visited page for capturing. Records with it the capture time and the source URL for a later reference. Capturing depth and quality can be customized for each capture operation such as: capture without styles, without scripts, edit before capture etc.
- Extensive capture – can capture web pages and files linked by the web page, multiple opened tabs – depth of capture such as 1st level or 2nd level or 3rd level links can also be defined for each capture operation separately.
- Data management – captured data can be organized into folders and stored in a tree structure, as easy as managing bookmarks. All data are stored locally, and is easy to locate the source directory or files in the file system. In addition, scrapbook structure and its references are stored as an .rdf file, which can be easily reused in external knowledge base for further analysis
- Search – data can be searched with a built-in fulltext search engine
- Editing – all captured resources can be edited by adding highlights, comments, annotations , links to external sources or other internally captured sources or even by editing the source HTML for the captured pages.
- Take notes – users can create local files or create note pages and edit them as easy as editing web pages. It is possible to define various HTML templates for various types of notes.

- Input and output data – the plug-in allows to combine multiple captured items or notes into a single one, it can generate a HTML tree list and make a static scrapbook site. In addition, multiple scrapbooks can be defined without interfering each other. Import and export of data items is possible in several formats.

Short overview on how above listed features fulfil the envisioned Digital Scrapbook approach is given in the Table 2 (note that the table does not contain full description of all Scrapbook-X features):

Digital Scrapbook feature	**Scrapbook-X support**
Web page capture (whole page)	*Capture page* ✓ Open the web page ✓ Right-click anywhere on the page (or select from the Scrapbook-X Menu in the browser menu bar) ✓ Capture page
Whole web page capture inclusive linked material	*Capture page as…* ✓ Open the web page ✓ Right-click anywhere on the page (or select from the Scrapbook-X Menu in the browser menu bar) ✓ Capture page as ...
Capture partial content (selection) from the web page	*Capture selection* or *Capture selection as*… ✓ Open the web page ✓ Select part of the web page to capture ✓ Right-click (or select from the Scrapbook-X Menu in the browser menu bar) ✓ Capture selection (or Capture selection as ..)
Annotate (comment)	*Create a Freenote* ✓ Open the captured page in the Scrapbook-X Sidebar ✓ Make a selection to be annotated ✓ In the footer of the browser window select the Annotation Tool (Lead pencil icon) ✓ Choose Create Freenote ✓ Enter the annotation content (text, link, etc.) The Freenote will be created directly under the selection. Note that this feature is not very stable for all selections.
Annotate (inline annotation)	*Add inline annotation to selection* ✓ Open the captured page in the Scrapbook-X Sidebar

	✓ Make a selection to be annotated ✓ In the footer of the browser window select the Annotation Tool (Lead Pencil icon) ✓ Choose Add inline annotation to selection ✓ Enter the text as annotation. Inline annotation is visible as "link" for the selected content. Hover over this link will display the text entered as annotation. Clicking on the link will open a dialog for editing of the annotated text.
Annotate (highlight)	*Highlighter Tool* ✓ Open the captured page in the Scrapbook-X Sidebar ✓ Make a selection to be highlighted ✓ Select the highlighter tool (Pencil icon) in the footer of the browser window and from its menu the highlight style Tipp: the last style used is selected by default, thus clicking on the pencil icon only is sufficient for the same highlight style
Manipulate Delete content /annotations/tags in captured sources	✓ Open the captured page in the Scrapbook-X Sidebar ✓ Make a selection to be deleted ✓ In the footer of the browser window select the Eraser Tool (Rubber icon) ✓ Choose the appropriate operation from the Eraser tool menu (delete all script tags, delete all iFrame tags, clear all scrapbook notes, clear all scrapbook notes in selection, remove the selection) Tipp: select content and click directly on the Rubber icon (do not select from menu) will remove the selection.
Manipulate Delete DOM area	✓ Open the captured page in the Scrapbook-X Sidebar ✓ In the footer of the browser window activate the DOM Eraser Tool (DOM icon) ✓ Hover over the area to remove from DOM (automatic highlight of current DOM area during hover) ✓ Click to remove the highlighted DOM area Tipp: DOM Eraser is enabled as well with keyboard shortcuts once DOM area is selected by moving the mouse (not clicking!). One can always invoke the shortcuts overview with "h" stroke when DOM Eraser is active.
Combine (view only)	✓ Select a folder in the Scrapbook-X Sidebar ✓ Right click on the folder and choose Combined view A combined view of all captured sources saved in selected folder is

	generated. This is only a view and is not a new resource.
Combine (create new resource)	✓ From the Scrapbook-X Sidebar invoke the *Tools* menu and choose *Combine Wizard* menu item ✓ From the left part of newly opened window choose resources which need to be combined in a new resource and drag them to the corresponding area on the right hand side ✓ Alternatively change the order in which the pages are combined ✓ Choose *Next* and wait for generating the *Preview* of the combined document ✓ Choose *Combine* to finalize the process New resource is generated from selected other resources, including in the source URL link to the combined content. Unfortunately this link is present only in the web content of the new resource, but not in the scrapbook RDF output[26].
View Original Source	✓ From the Scrapbook-X Sidebar navigate to captured resource ✓ Use the mouse right-click and *Open source URL* from the context menu Current browser view will be navigated to the original resource. When this feature is invoked for combined documents it will only open the source URL of the first document in the list. Tipp: alternatively, from the context menu (available on mouse right-click) and then the *Tools* submenu, user may capture the content again.
Link from Source to Scrapbook	 This feature is not implemented; in general, it will require access to a Source Server[27].
Search	✓ In the Scrapbook-X Sidebar a search criteria can be provided in the search field for fulltext search ✓ Alternatively, users may search by id, title, source URL or a comment (annotation) content
Publish as Single File	✓ In the Scrapbook-X Sidebar select a resource to publish ✓ User the mouse right-click to open the context menu ✓ Choose Create MAF menu option A single file in a MAF (Mozilla Archive Format) is generated, which can

[26] A contact to developer may resolve this issue
[27] See more in Towards a digital Scrapbook Environment section

	be viewed in a Firefox browser with the MAF Reader extension.

Table 2: An overview of the digital scrapbook features and the corresponding Scrapbook-X features

More information and up-to date user documentation can be found directly at the plug-in development page (see Table 1). The Appendix chapter in this document contains screenshots created during interaction with the Scrapbook-X.

Towards a Digital Scrapbook environment

Scrapbook-X plugin supports a significant subset of the features necessary to implement the Scrapbook approach. Alone it may be well applied as a "personal" note taking and source-capturing tool, effective for personal collection management of captured web resources. To support the digital scrapbook approach it however lacks more basic features, some of which are mentioned as follows:

- *Editorial support* – offer various layout templates for combined resources, such that the new scrapbooks could comply with more sophisticated presentation requirements. For example, users can simply drag-and-drop captured resources to a layout template of choice; potentially enable simple publishing workflow to distinguish between "publishable" and "non-publishable" parts of a combined scrapbook resource i.e. distinguish between simple comments which should be kept private and attributions which could be shared with other scrapbook readers; generate a list of "scrapbook bibliographical references" Full auditable trail to sources (provenance and smart metadata capture)
- *Share and publish as Open access* – the scrapbook content should be publicly available, as open access, and published in an appropriate manner i.e. assigned with a DOI and some bibliographic metadata, so that it can be shared with other users, disseminated via a repository, or included as a reference in the users' personal reference management system
- *Evolving knowledge base* – annotations and tagging of facts on top of a scrapbook resource evolve the knowledge base around a topic, for example, better visualization of scrapbook or external resources around a particular research question or a topic (see discussion below)
- Standards-based Interoperability of resources – resources should be saved in an interoperable formats, described with wide accepted standards, such as published ontologies in RDF or JSON-LD format, to leverage their reuse and enable development of new features
- *Cross-browser, cross-platform and cross-device support* – scrapbook component should operate agnostic to the operating system in use, file system in use, browser or device in use. Desktops are no longer only device of choice when it comes to web-based research activity neither is a single browser in use for an application. Thus the Scrapbook component must be built upfront with this cross-compatibility in mind. A good example is the recently announced approach to the Firefox Add-Ons development (Needham, 2015), which will certainly influence the Scrapbook Family of Add-ons as well.
- *Back-reference from origin web-sites to the scrapbook itself* – will enhance the scrapbook visibility, but also the visibility of resources included in any scrapbook. For this purpose, we may consider both technical and organizational approaches. For example, the Scrapbook tool informs the user about the inclusion of a resource into one or more Scrapbooks whenever user navigates

to that resource on the web. This local approach would be relatively easy[28] to solve within the Scrapbook tool and may slightly influence the speed of web-surfing. A real challenge would be to solve the back-referencing for a shared, multi-user Scrapbook platform – it raises both technical (too much client-server communication) and privacy issues (server practically gets a log of user browsing activity). Potential solution would be e.g. to generate an "index" of existing scrapbook sources and their origin URLs, and synchronizing it locally at researcher devices, thus operating similar to the local approach mentioned before. The index must be optimized and should not affect the response latency. This is of course, only a partial contribution to the back-referencing problem. In addition, with organizational actions such as agreements on including links to scrapbook resources at the origin URLs with relevant web publishers will be of great benefit for both scrapbook and original resource.

- *Authentication and authorization* –these mechanisms must be supported for several reasons: only authorized users can collaboratively edit a single scrapbook resource, the provenance of created comments, annotations or user-generated data is automatically ensured, a researcher may feel more comfortable working in "single-user" mode for some time, and later open the scrapbook resource for annotations, comments and tagging by other users. Moreover, even if the scrapbook is completely public from the start, a system that supports authorization and has clearly implemented user roles is rather trusted and reliable for end users.

More advanced features should additionally enable networked sources, services and data, available both centrally and locally, with the goal to improve sharing of scrapbook resources as open access, to enhance collaboration and on the other hand side, to allow scholars to work locally - in an own, personalized and private environment.

The scrapbook environment should be well integrated with the standard research processes such as: notes taking, collection, classification and prioritization of sources, writing hypotheses, writing "papers" etc. It should deliver products which are both outputs and sources for further scholarly works. It should be language and scholarly domain polyglot in order to support specific research questions. It should be integrated with search engines and natural language processing tools, to boost the discovery of resources and at the same time allow for evolution of user-generated knowledge base (KB); the user generated KB, although inherently populated during the process of annotation and linking of sources with specialized facts, should be available for reuse and repurpose through reasoning, inferencing and further interlinking with domain-specific or global knowledge bases such as DbPedia[29], BabelNet[30], GeoNames[31] and others[32].

Envisioning the digital scrapbook implementation as an isolated application available on the web, or as single self-sufficient tool would be an approach which may provide some results on short-term basis. Looking from longer term perspective, such a tool will be probably outdated at the moment it is deemed to be finalized: the Web is in continuous update, new technologies, tools, data, formats, services and usages emerge, while many of existing are dismissed and disappear under the Hammer

[28] We should however be aware of the complexity imposed by existence of multiple URLs even for resources within a same domain i.e. sometimes a resource can be reached with a "pretty URL"; sometimes same resource is resolved with a URL which contains more parameters such as language, format, sessionId etc.
[29] See http://wiki.dbpedia.org/about
[30] See http://babelnet.org/about
[31] See http://www.geonames.org/
[32] See LOD Cloud at http://lod-cloud.net/

of Innovation. Thus the Scrapbook implementation needs to be a long term continuous software development effort to sustain meaningful.

The Scrapbook (user) experience

Previous discussion addressed a set of features and aspects which need to be carefully considered from the very beginning of the Scrapbook design. Some may be considered as foundational, others as "nice-to-have" or will be never integrated in a Scrapbook solution. New features may emerge from already available or newly developed services and toolkits and could prove as necessary to humanities scholars, thus must be integrated within the Scrapbook environment as well.

Regardless if it is a simple Firefox Add-On, or implemented as a more robust platform, it is clear that the Scrapbook environment should adhere to the basic rules for user acceptance: simple and user-friendly enough, yet feature-rich and integrated with tools researchers are comfortable to work with (such as browsers, search engines, text or table editors tools), "facilitating their activities, rather than determining how these should be performed" (Edmond, Bulatovic, & O'Connor, 2015, pp. 13-14). The bar for good user experience for a research and office productivity tools has recently been set very high by the user experiences of current Office products (as. e.g. Microsoft Word), scientific mathematical tools (as e.g. Mathematica) and graphical user interfaces of popular mobile device operating systems and related apps (e.g. iOS or Android).

Implementing a Scrapbook platform with a vision of a preferred tool to publish combined resources and ensure open access to digital humanities, is not and cannot be seen as a mere selection from a range of Javascript libraries packed with beautifully designed user interface, supporting HTML 5[33] recommendations. Ensuring satisfied users and great user experience[34] (UX) with the Scrapbook environment is challenging and comprises activities additional to the web-design. These have to be addressed continuously, start before any development and continue as long as the product is in use.

The user experience (UX), as a quality experienced by the user during the interaction with the software is dynamic, changes and evolves as users learns more about the software and experience usage of other similar products, or increment in requirements to perform their activities. There are numerous studies and research around this topic, but, although it is not newly emerged, only in recent years it starts to be more intensively applied in scholarly tools and applications. UX methods and research can be grouped around following aspects[35]:

- Utility – how useful are the features offered by the software and do these fit its intended purpose? Are Scrapbook features helping users to perform their research, or are yet another unnecessary distortion?
- Usability – can user perform his activities within the Scrapbook in an easy and efficient manner? Can users capture and organize resources in a desired layout? Can they quickly find them?

[33] See https://en.wikipedia.org/wiki/HTML5
[34] See https://en.wikipedia.org/wiki/User_experience
[35] See https://en.wikipedia.org/wiki/User_experience_evaluation

- Aesthetics – is the software visually attractive to the user? Are colours and fonts carefully selected and pleasant for users? Can they be used safely in the presentation of valuable research outputs?
- Stimulation – can the software inspire users for further contributions i.e. would a Scrapbook resource become recommended method for innovative publishing? Has it become "native" item in the user tools library, thus users are happy to use it whenever possible?
- Value – how important is the Scrapbook environment and its products to the end user? Can a Scrapbook resource be valuable to own career path?
- Fulfilment of user expectations: if a user invests time to learn the software and creates valuable (to her) scrapbook resources, how valuable these resources would be for the others? Would a Scrapbook resource become a "brand" and be accepted as any other journal article?

Successful Scrapbook environment requires involvement of researchers from very first steps of development in close collaboration with technical team and user experienced members – and has to establish continuous software development and continuous UX testing and feedback in both directions.

Conclusion

Discussions about the features and UX aspects only contribute to the notion that the Scrapbook environment shall be sustainable and trustworthy service, surrounded with supporting activities around researchers, such as helpdesk, training and promotional material, user experience evaluation, stabile connections with the end users etc.

The Scrapbook environment shall in addition be alert for unprecedented use. This can technically be ensured by a design approach and selection of technologies that are in harmony with the Web and its dynamics, for example, by employing java-script based components and integration of external services.

Building a Scrapbook environment for researchers, inclusive for new services, resources and features, which is natural to the Web itself does not finish with the software development end. To fulfil its purpose, it needs to be delivered as a service and actually used by researchers: it is continuous, time-consuming and resource-heavy activity which requires several person years for software development and multi-person team possessing expertise in digital humanities, software development and architecture, web science and user experience methods and evaluation. Aiming at a relevant user experience and judging from our experience we recommend to calculate with an investment of approximately 10 person years for the initial productive version of the software, which includes UX engineering, 2 full time equivalents for continuous software and UX improvements and 2 full time equivalents for supporting a larger group of installations including technical and basic user support, with a beta version of the software delivered at the half of the development period.

References

Berners-Lee, T. (2006). *Design Issues - Linked Data.* (W. W. Consortium, Ed.) Retrieved 10 08, 2015, from http://www.w3.org/DesignIssues/LinkedData.html

Cambria, E., & B.White. (2014). Jumping NLP Curves. (IEEE, Ed.) *IEEE Computational Intelligence Magazine, 9*(2), pp. 48-57.

Edmond, J., Bulatovic, N., & O'Connor, A. (2015). The Taste of "Data Soup" and the Creation of a Pipeline for Transnational Historical Research. *Journal of the Japanese Association for Digital Humanities, 1*(1), pp. 107-122.

Malcolm D., H., & Renn, J. (2012). Toward an Epistemic Web. In J. (. Renn, *The Globalization of Knowledge In History* (pp. 821-838). Berlin, Germany: Edition Open Access.

Maurer, T. (2014). Retrieved 10 08, 2015, from This is Why OneNote is Awesome: http://www.thomasmaurer.ch/2014/02/this-is-why-onenote-is-awesome/

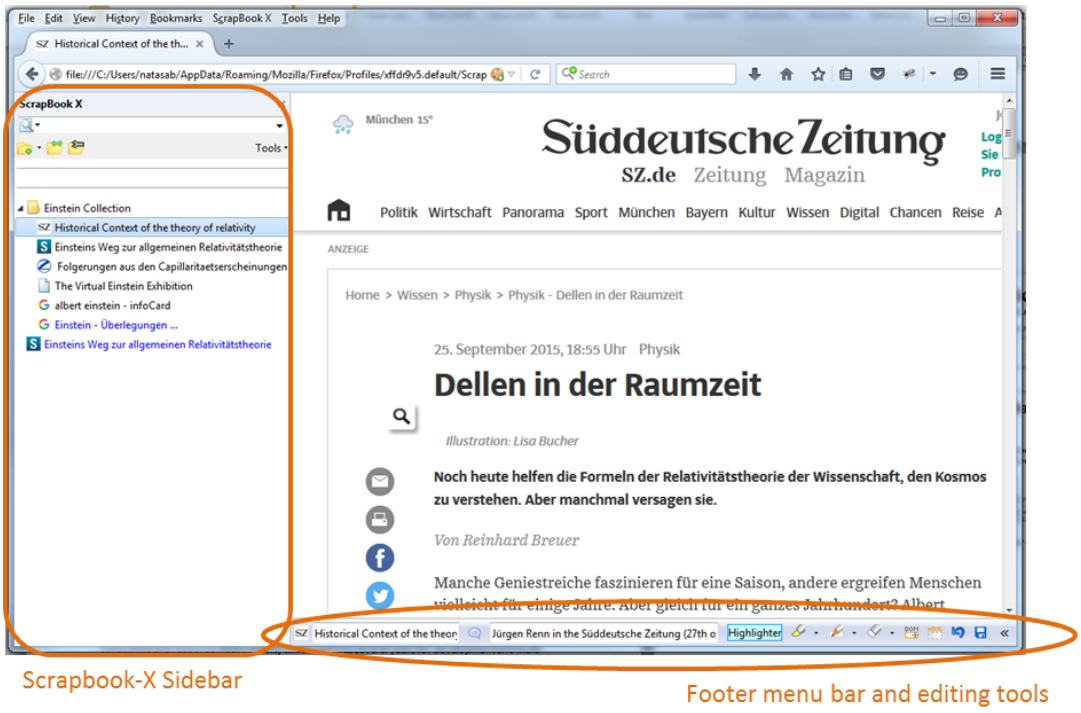
Needham, K. (2015). *The future of developing Firefox Add-ons*. Retrieved 10 08, 2015, from Mozilla Add-Ons Blog: https://blog.mozilla.org/addons/2015/08/21/the-future-of-developing-firefox-add-ons/

Pawliczek, A. (2014). *Building up a Research Infrastructure on the First World War across Borders.* Retrieved 10 08, 2015, from http://www.apex-project.eu/index.php/en/articles/180-building-up-a-research-infrastructure-on-the-first-world-war-across-borders

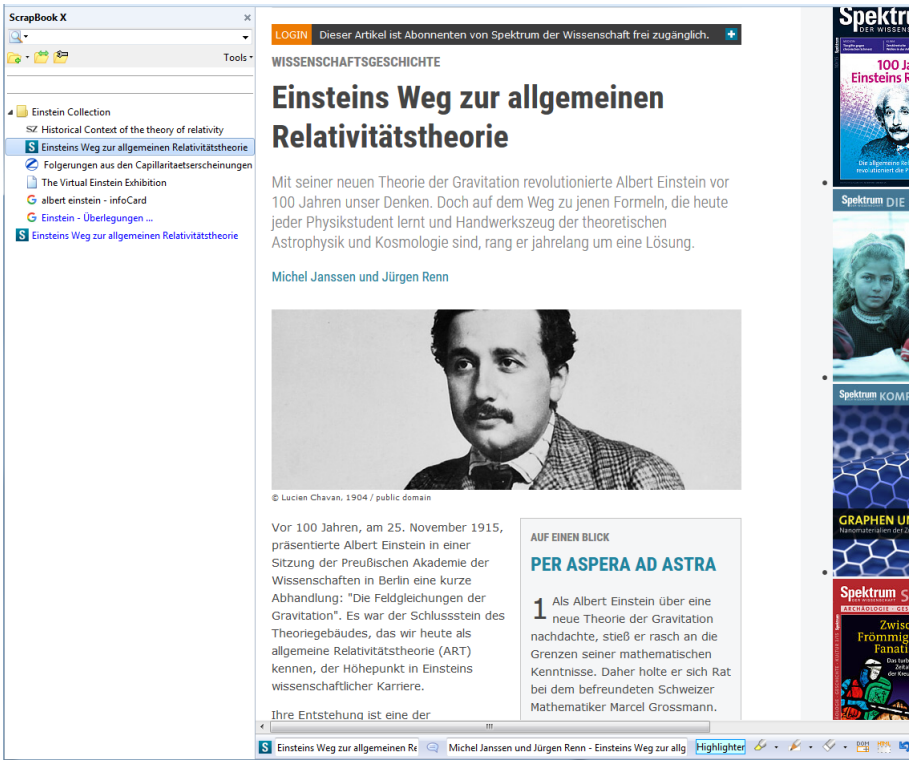
Renn, J. (2014). Beyond editions: historical sources in the digital age. In M. Stolz, Y.-C. Chen, & (Eds.), *Internationalität und Interdisziplinarität der Editionswissenschaft* (Vol. 38, pp. 9-28). Berlin, Germany: De Gruyter.

Renn, J. (2015). Fiesole Collection Development Retreats. *The value of collections for researchers in the humanities*. Berlin: Casalini libri. Retrieved from http://www.casalini.it/retreat/retreat_2015.html

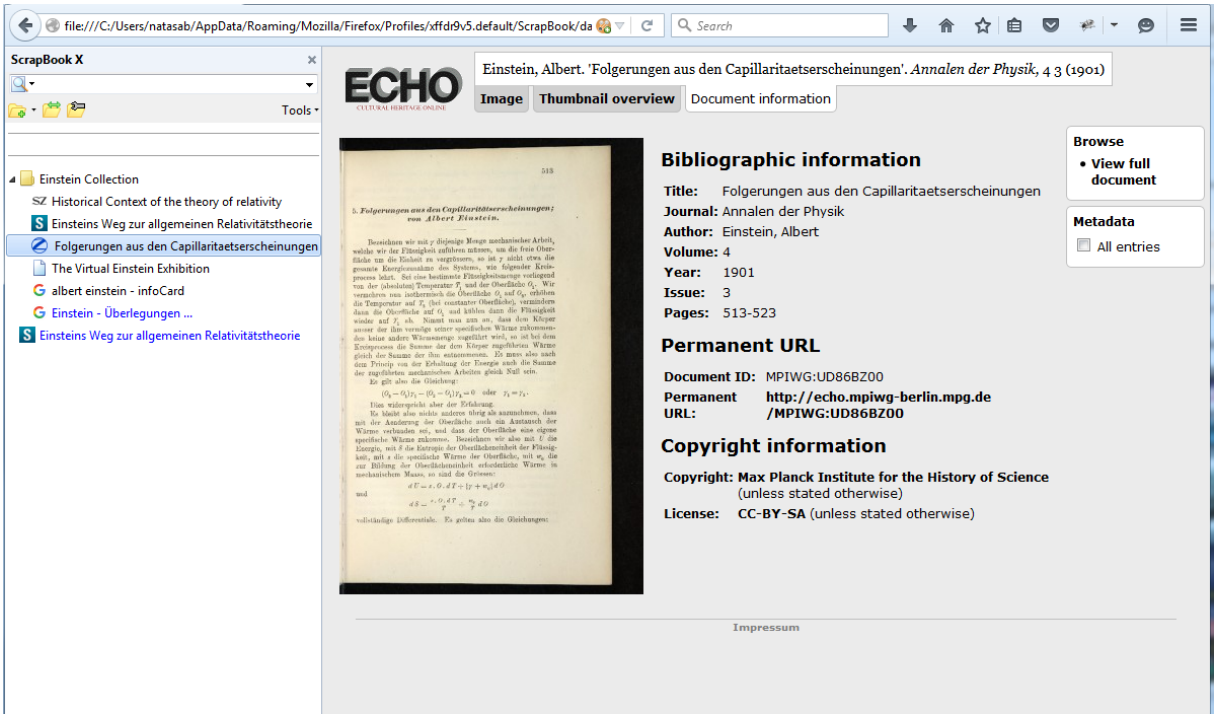
Appendix



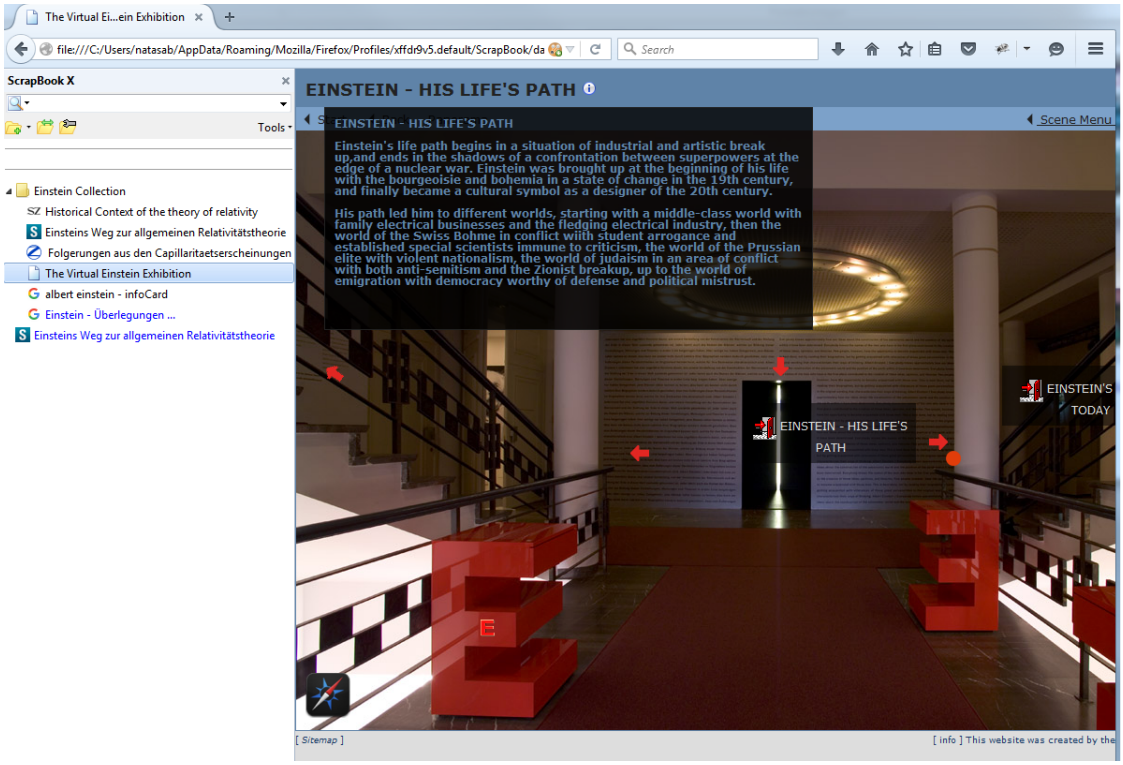
Standard view of the Scrapbook-X Add-on as displayed in the browser. Current browser window displays an Article about Einstein, captured from *Süddeutsche Zeitung*, http://www.sueddeutsche.de/wissen/physik-dellen-in-der-raumzeit-1.2664283



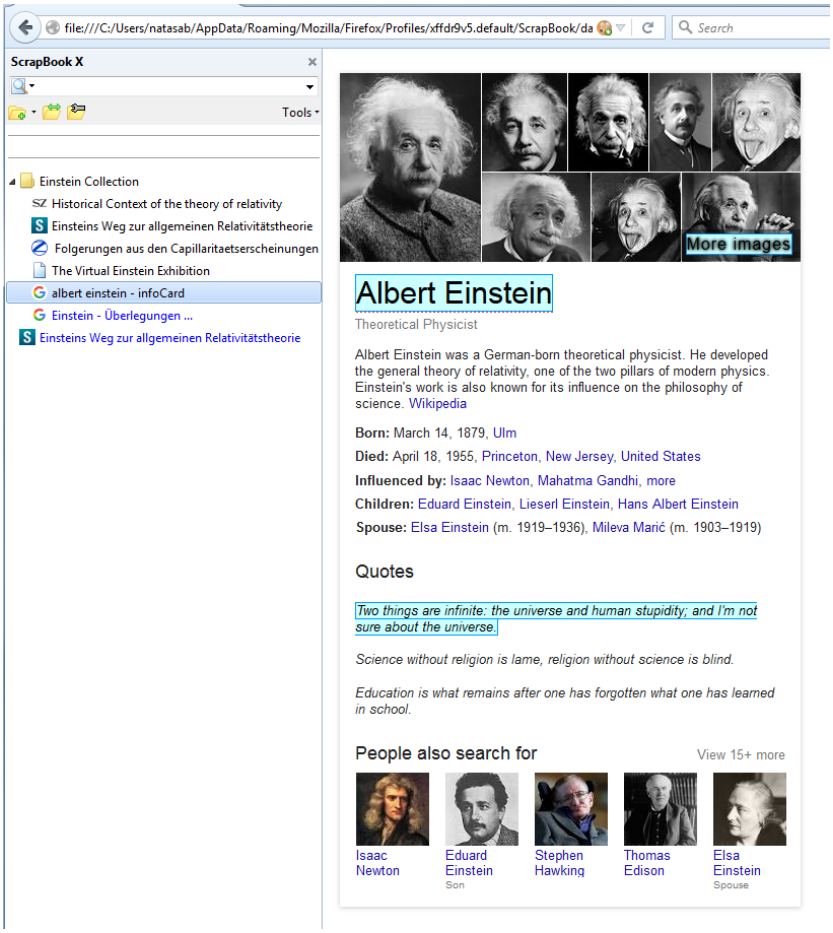
An Article on Einstein, captured from *Spektrum.de*, http://www.spektrum.de/magazin/einsteins-weg-zur-allgemeinen-relativitaetstheorie/1362267



An ECHO collection resource from Albert Einstein, http://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/permanent/echo/einstein/annalen/Einst_Folge_de_1901/index.meta&viewMode=index



One view from the Virtual Einstein exhibition, http://einstein-virtuell.mpiwg-berlin.mpg.de/VEA/SC1615465739_en.html?previousScene=SC879771616_en.html



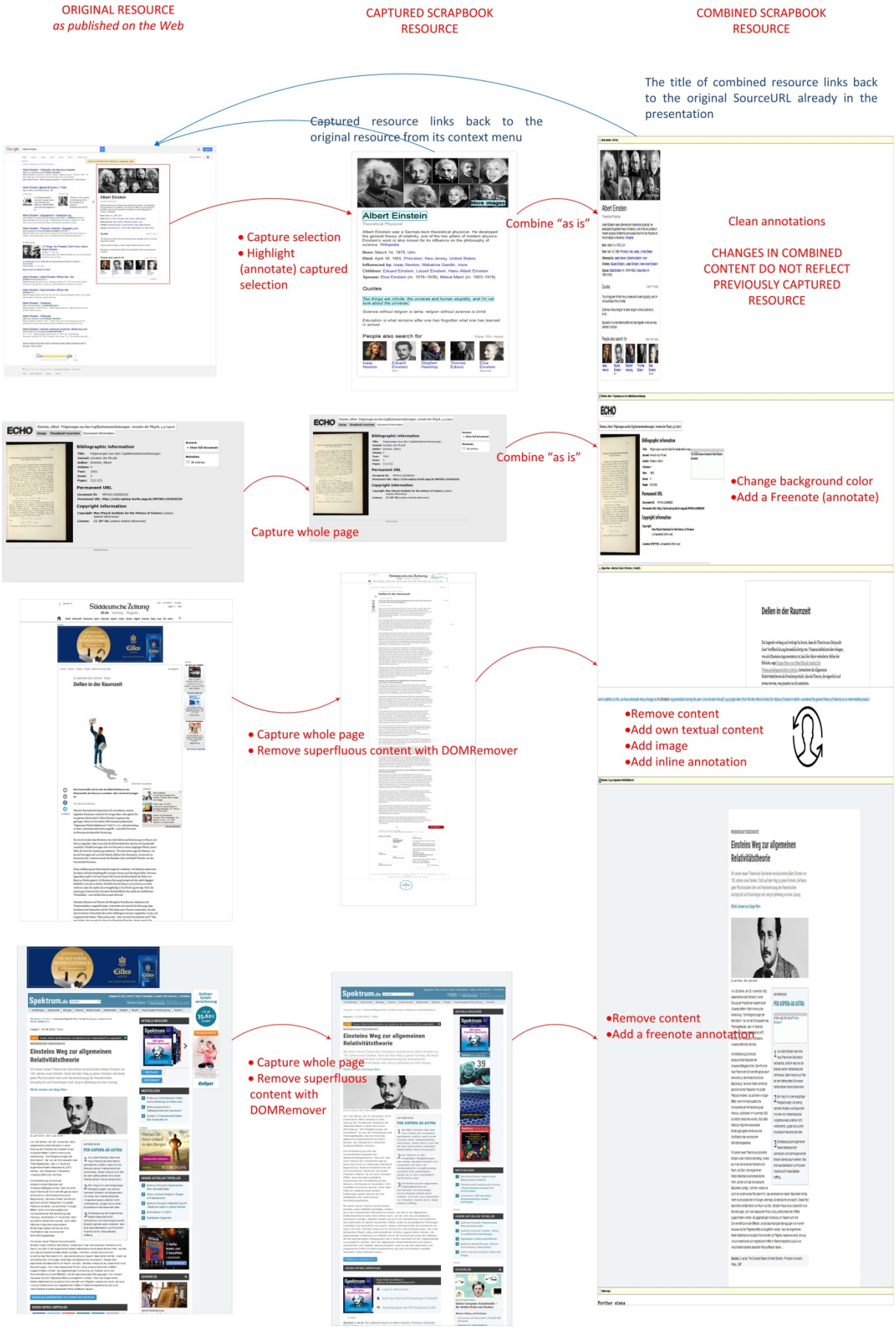
Selection of a page content generated by Google search on “Albert Einstein”. The content generated by Google knowledge graph has been selected and captured. Captured data have been annotated (see coloured highlights on the right-hand side in the screenshot above).

A new combined resource, named “Einstein-Überlegungen…” has been created, with following actions applied:

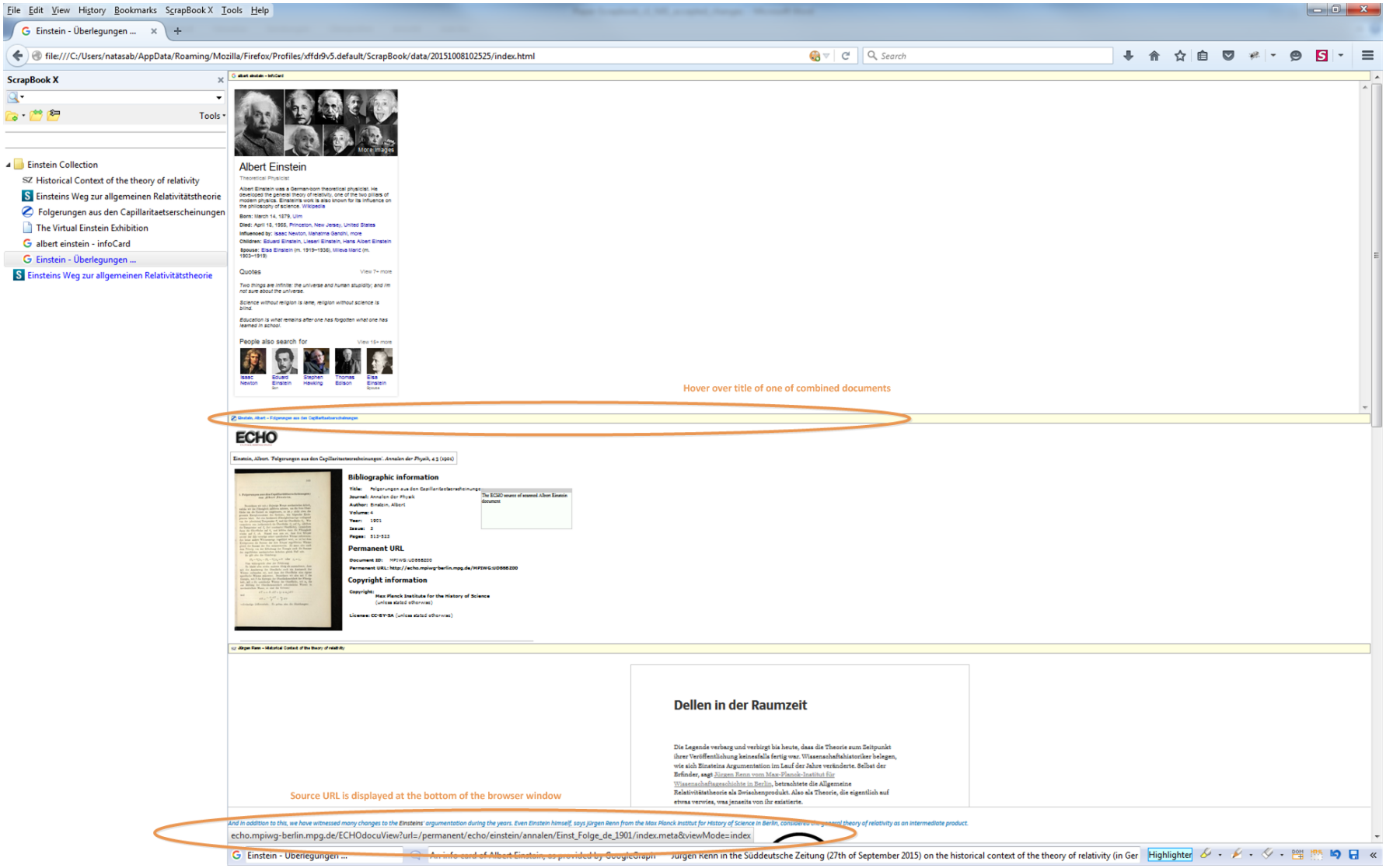
1) All resources captured have been combined in a new document. The sections of the newly combined document have been edited as follows:
    - Echo Collection Resource (edited background)
    - Albert Einstein partially captured search result (highlights are removed)
    - The article from Süddeutsche Zeitung has been reduced only to one paragraph, all other graphical elements have been removed
    - The Article from Spektrum.de has been reduced to contain only the text from the Einstein Article
2) For some resources additional annotations have been provided:
    - Echo Collection Resource – a freenote comment has been added
    - To the remained content of the article from Süddeutsche Zeitung additional content has been added in a form of text, local image has been inserted, and an “inline” annotation has been added, which links a part of the text to external resource on the Web

3) To the article of Spektrum.de additional freenote comment has been added. The name used in this freenote comment has been linked in addition to a Wikipedia Entry.
4) A new standalone note has been created, which contains “Further steps” as a reminder.
5) Newly created note has been combined with the previously combined document from step 1 and has been added at the end of that document
6) The note has been deleted after it was combined in step 4. After this step, it only resides in the combined document from step 4 and not as a standalone note.

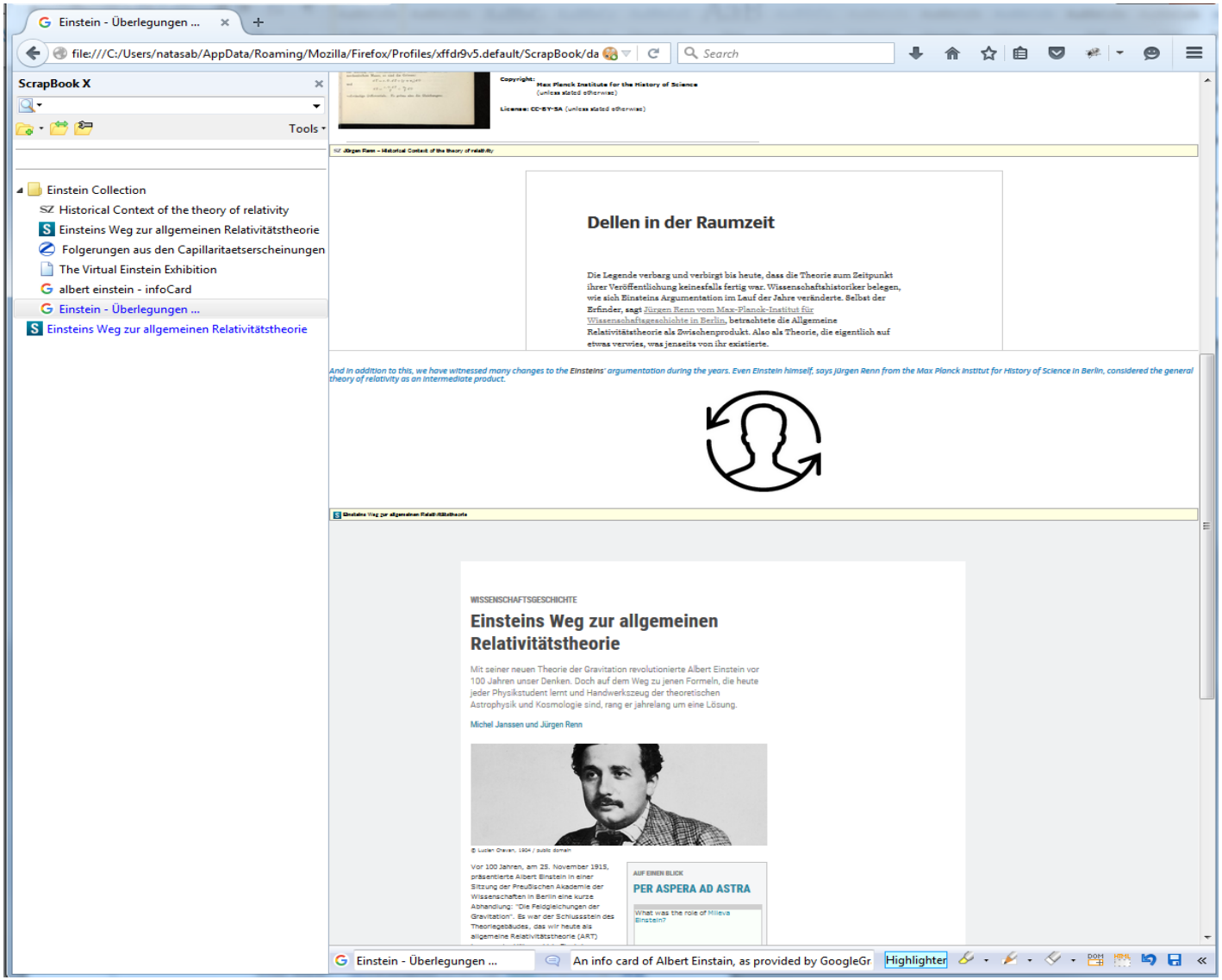
Resulting document after steps 1, 2 and 3 is displayed with several screenshots in the remaining of the document.



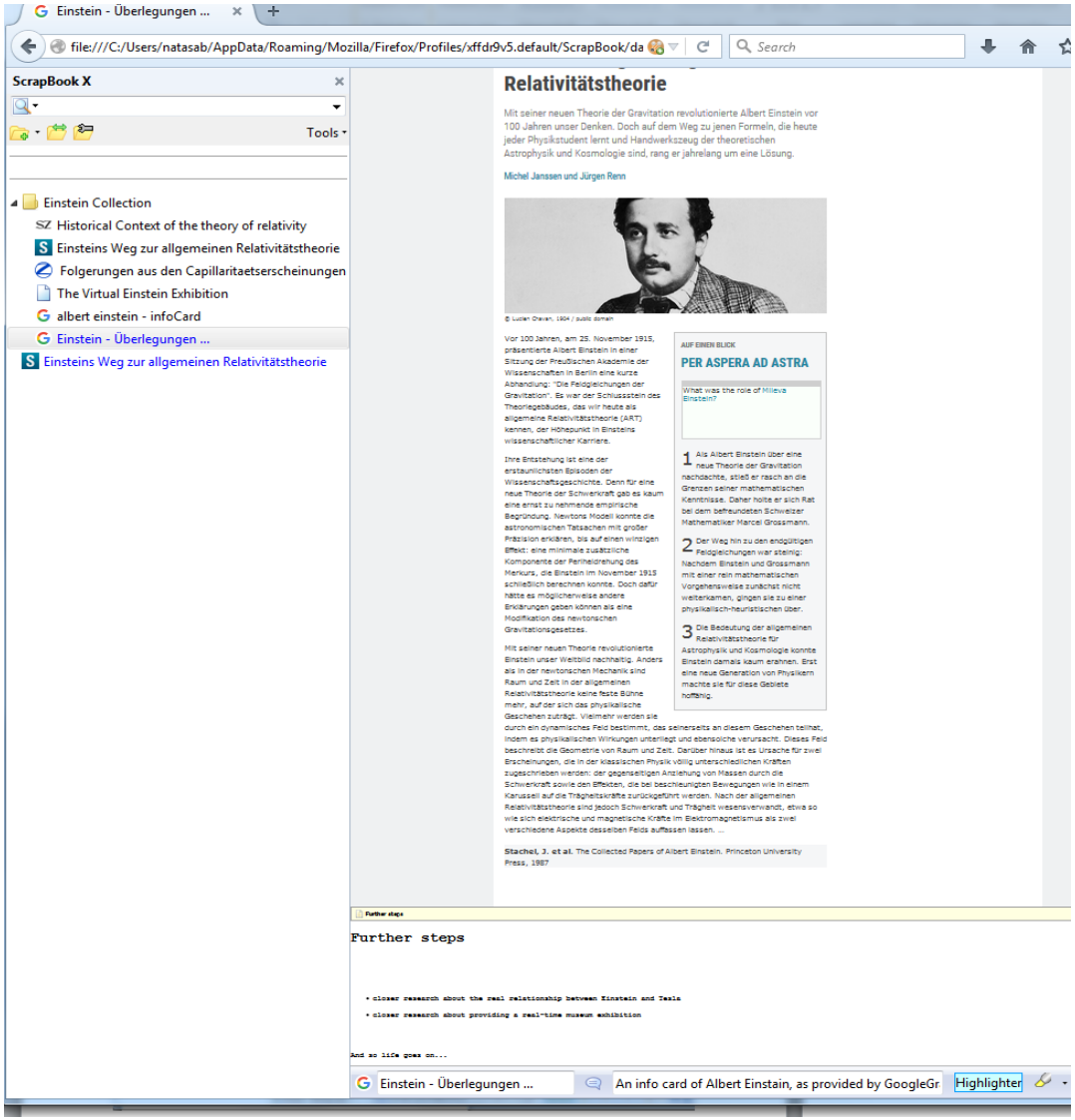
Web resources, scrapbook resources and combined scrapbook resource



Combined document – Part 1



Combined document – Part 2



Combined document – Part 3

MAX-PLANCK-INSTITUT FÜR WISSENSCHAFTSGESCHICHTE

Max Planck Institute for the History of Science

Preprints since 2013 (a full list can be found at our website)

437 Jürgen Renn **Schrödinger and the Genesis of Wave Mechanics**

438 Pietro Daniel Omodeo **L'*iter* europeo del matematico e medico scozzese Duncan Liddel**

439 Irina Tupikova & Klaus Geus **The Circumference of the Earth and Ptolemy's World Map**

440 Pietro Daniel Omodeo und Jürgen Renn **Das Prinzip Kontingenz in der Naturwissenschaft der Renaissance**

441 Horst Kant und Jürgen Renn **Eine utopische Episode – Carl Friedrich von Weizsäcker in den Netzwerken der Max-Planck-Gesellschaft**

442 William G. Boltz and Matthias Schemmel **The Language of 'Knowledge' and 'Space' in the Later Mohist Canon** (TOPOI – Towards a Historical Epistemology of Space)

443 Stefano Bordoni **Looking for a Rational Thermodynamics in the late XIX century**

444 Sonja Brentjes and Jürgen Renn **The Arabic Transmission of Knowledge on the Balance**

445 Horst Nowacki **Archimedes and Ship Design**

446 Matthias Schemmel **Elements of a Historical Epistemology of Space** (TOPOI – Towards a Historical Epistemology of Space)

447 Martin Thiering and Wulf Schiefenhövel **Spatial Concepts in Non-Literate Societies: Language and Practice in Eipo and Dene Chipewyan** (TOPOI – Towards a Historical Epistemology of Space)

448 Jürgen Renn **Einstein as a Missionary of Science**

449 Hubert Laitko **Der Ambivalenzbegriff in Carl Friedrich von Weizsäckers Starnberger Institutskonzept**

450 Stefano Bordoni **When Historiography met Epistemology.** Duhem's early philosophy of science in context

451 Renate Wahsner **Tausch – Allgemeines – Ontologie** oder **Das Auseinanderlegen des Konkreten und seine Aufhebung**

452 Jens Høyrup **Algebra in Cuneiform.** Introduction to an Old Babylonian Geometrical Technique

453 Horst Nowacki **Zur Vorgeschichte des Schiffbauversuchswesens**

454 Klaus Geus and Mark Geller (eds.) **Esoteric Knowledge in Antiquity** (TOPOI – Dahlem Seminar for the History of Ancient Sciences Vol. II)

455 Carola Sachse **Grundlagenforschung. Zur Historisierung eines wissenschaftspolitischen Ordnungsprinzips am Beispiel der Max-Planck-Gesellschaft (1945–1970)**

456 David E. Rowe and Robert Schulmann **General Relativity in the Context of Weimar Culture**

457 F. Jamil Ragep **From Tūn to Turun: The Twists and Turns of the Ṭūsī-Couple**

458 Pietro Daniel Omodeo **Efemeridi e critica all'astrologia tra filosofia naturale ed etica: La contesa tra Benedetti e Altavilla nel tardo Rinascimento torinese**

459 Simone Mammola **Il problema della grandezza della terra e dell'acqua negli scritti di Alessandro Piccolomini, Antonio Berga e G. B. Benedetti e la progressiva dissoluzione della cosmologia delle sfere elementari nel secondo '500**

460 Stefano Bordoni **Unexpected Convergence between Science and Philosophy: A debate on determinism in France around 1880**

461 Angelo Baracca ***Subalternity vs. Hegemony* – Cuba's Unique Way of Overcoming Subalternity through the Development of Science**

462 Eric Hounshell & Daniel Midena **"Historicizing Big Data" Conference, MPIWG, October 31 – November 2, 2013** Report

463 Dieter Suisky **Emilie Du Châtelet und Leonhard Euler über die Rolle von Hypothesen.** Zur nach-Newtonschen Entwicklung der Methodologie

464 Irina Tupikova **Ptolemy's Circumference of the Earth** (TOPOI – Towards a Historical Epistemology of Space)

465 Irina Tupikova, Matthias Schemmel, Klaus Geus **Travelling along the Silk Road: A new interpretation of Ptolemy's coordinates**

466 Fernando Vidal and Nélia Dias **The Endangerment Sensibility**

467 Carl H. Meyer & Günter Schwarz **The Theory of Nuclear Explosives That Heisenberg Did not Present to the German Military**

468 William G. Boltz and Matthias Schemmel **Theoretical Reflections on Elementary Actions and Instrumental Practices: The Example of the *Mohist Canon*** (TOPOI – Towards a Historical Epistemology of Space)

469 Dominic Olariu **The Misfortune of Philippus de Lignamine's Herbal or New Research Perspectives in Herbal Illustrations From an Iconological Point of View**

470 Fidel Castro Díaz-Balart **On the Development of Nuclear Physics in Cuba**

471 Manfred D. Laubichler and Jürgen Renn **Extended Evolution**

472 John R. R. Christie **Chemistry through the 'Two Revolutions': Chemical Glasgow and its Chemical Entrepreneurs, 1760-1860**

473 Christoph Lehner, Helge Wendt **Mechanik in der *Querelle des Anciens et des Modernes***